

Noise statistics in a fast digital radio receiver: the Bedlam backend for the Parkes Radio Telescope

J.D. Bray · R.D. Ekers · P. Roberts

Received: date / Accepted: date

Abstract The digital record of the voltage in a radio telescope receiver, after frequency conversion and sampling at a finite rate, is not a perfect representation of the original analog signal. To detect and characterise a transient event with a duration comparable to the inverse bandwidth it is necessary to compensate for these effects, which modifies the statistics of the signal, making it difficult to determine the significance of a potential detection. We present an analysis of these modified statistics and demonstrate them with experimental results from Bedlam, a new digital backend for the Parkes radio telescope.

Keywords Coherent pulse detection · Gaussian noise statistics · Digital radio receivers · ADC non-linearity

1 Introduction

Typical radio astronomy applications are not sensitive to transitory extreme fluctuations in the signal voltage from the telescope receiver. For example, when measuring the spectral intensity of a static astronomical object, mere single-bit sampling of the voltage achieves a minimum of 64% of the signal-to-noise ratio for the ideal many-bit case, even though the ability to measure the magnitude of a single sample is entirely lost (Weinreb, 1963, sec. III). However, in applications in which a single high-magnitude sample is observationally meaningful — implying a signal with time structure on the scale of the inverse bandwidth — it becomes necessary to examine the single-sample behaviour of the telescope receiver, and to develop an understanding of its statistics.

J.D. Bray
School of Chemistry & Physics, University of Adelaide, Australia
E-mail: justin.bray@adelaide.edu.au

J.D. Bray · R.D. Ekers · P. Roberts
CSIRO Astronomy & Space Science, Epping, Australia

One such application is the search for nanosecond-scale broad-band radio pulses from neutrino-induced particle cascades in the Moon, currently being conducted with the WSRT (Buitink et al., 2010), EVLA (Jaeger et al., 2010) and ATCA (James et al., 2010) telescopes. These pulses are inherently very narrow in time, with a detectable width approximately equal to the inverse bandwidth resulting from band-limiting in the radio receiver. A single such pulse would appear in the voltage recorded by the receiver as a transitory excursion — possibly only a single sample — beyond the statistical bounds of the background Gaussian noise, here dominated by thermal radiation from the Moon. If such an event were detected, and a possible origin as radio frequency interference (RFI) were excluded, it might indicate the registration of an ultra-high energy neutrino, with significant implications for cosmic ray astrophysics (Beresinsky and Zatsepin, 1969). The confidence of the detection would depend on the expected rate of false positives from noise in the receiver, which we address in this paper.

Radio receivers typically convert the incoming signal from its original radio frequency to a lower intermediate frequency by mixing it with the single-frequency output from a local oscillator. This process affects time-averaged properties of the signal, such as the spectral intensity, in a predictable fashion, but the effect on the shape and height of an individual peak will depend on its phase relative to the local oscillator. Since this phase is usually unknown, the original signal cannot be unambiguously reconstructed; to search for peaks in the original signal, it is necessary to reconstruct all of its possible states, searching in phase for the maximum peak height. In section 2 we illustrate this effect, and show that the signal envelope defined by Longuet-Higgins (1957) reconstructs the maximum possible peak height.

The fraction of digital samples above some detection threshold can be trivially calculated with Gaussian statistics. However, a peak in the original analog signal will generally lie between two sampled points, neither of which will record its full magnitude, as first pointed out in this context by James et al. (2010). The values of the analog signal between the sampled points, including any such peaks, can be reconstructed through convolution with a sinc function as a consequence of the Nyquist-Shannon sampling theorem. The number of expected excursions beyond some threshold in this interpolated signal was derived by Rice (1944, 1945); in section 3 we explain the application of this result and extend it to find the excursion rate for the signal envelope.

We test the above results through simulation and experiment. In section 4 we perform Monte Carlo simulations of the excursion rates for noise with various spectra, finding them to be consistent with the analytic results. In section 5 we present the results of an experiment in which a noise source was connected to “Bedlam” — a new digital backend for the Parkes radio telescope — which recorded extreme fluctuations in the signal voltage for a period of 20 days. Bedlam was designed and built for LUNASKA (Lunar Ultra-high energy Neutrino Astrophysics with the Square Kilometre Array), a neutrino detection project of the type described above (Bray et al., 2012). The experiment described here allows us to test its consistency with analytic and simulated models of the noise behaviour, and to calibrate any variation.

2 Frequency mixing

Shifting a signal in frequency by mixing it with a local oscillator signal, as in a typical radio receiver, also changes the phase of the signal. To find a possible peak in the original signal, it is necessary to search through the possible values of this phase to find the maximum peak height. Here we illustrate this effect, and show that the phase-searched maximum signal can be represented by an analytic envelope of the signal.

2.1 Origin of the phase ambiguity

The mixing stage in a radio receiver can be represented as a multiplication between the original radio frequency signal $f_{\text{RF}}(t)$ and a local oscillator signal $f_{\text{LO}}(t)$, equivalent to a convolution $F_{\text{RF}}(\nu) * F_{\text{LO}}(\nu)$ in the frequency domain. The local oscillator and its Fourier transform $F_{\text{LO}}(\nu)$ are given by

$$\begin{aligned} f_{\text{LO}}(t) &= \cos(2\pi \nu_{\text{LO}} t + \phi) \\ F_{\text{LO}}(\nu) &= \frac{1}{2} e^{i\phi} \delta(\nu - \nu_{\text{LO}}) + \frac{1}{2} e^{-i\phi} \delta(\nu + \nu_{\text{LO}}) \end{aligned} \quad (1)$$

where ϕ describes the phase of the local oscillator relative to the signal. After subsequent bandpass filtering to obtain the low-frequency component, the resulting intermediate frequency signal $f_{\text{IF}}(t)$ is described, in the frequency domain, as:

$$F_{\text{IF}}(\nu) = \frac{1}{2} e^{i\phi} F_{\text{RF}}(\nu - \nu_{\text{LO}}) + \frac{1}{2} e^{-i\phi} F_{\text{RF}}(\nu + \nu_{\text{LO}}) \quad |\nu| < \nu_{\text{LO}}. \quad (2)$$

For a low-side local oscillator (i.e. ν_{LO} is less than all frequencies ν present in the radio frequency signal), this becomes

$$F_{\text{IF}}(\nu) = \frac{1}{2} e^{-i\phi \text{sign}(\nu)} \begin{cases} F_{\text{RF}}(\nu + \nu_{\text{LO}}) & 0 < \nu < \nu_{\text{LO}} \\ F_{\text{RF}}(\nu - \nu_{\text{LO}}) & -\nu_{\text{LO}} < \nu < 0 \end{cases} \quad (3)$$

showing that the intermediate frequency signal depends on the phase factor $e^{-i\phi \text{sign}(\nu)}$. The expression for a high-side local oscillator is similar, but with the sign of ϕ reversed; additional mixing stages add additional phases ϕ_1 , ϕ_2 , etc. If the phase of the local oscillator is unknown, the effect in these cases is the same: the intermediate frequency signal includes an unknown phase ϕ in the range $0 \rightarrow 2\pi$.

2.2 Phase searching and the signal envelope

The effect in the time domain of introducing a variable phase ϕ can be to decrease the peak amplitude of the pulse, as illustrated in figure 1. As ϕ is unknown, it is not possible to unambiguously reconstruct the original signal. However, given a post-mixing signal $f(t) = f_{\text{IF}}(t)$, it is possible to reconstruct the different possible states of the original signal for different values of ϕ . An excursion beyond some threshold

in the original signal must appear as a peak of at least that height in one or more of these reconstructed signals. We define $h(t, \theta)$ as the reconstructed form of $f(t)$ with maximum peak height; in terms of their respective Fourier transforms, $H(v, \theta)$ and $F(v)$, it is

$$H(v, \theta) = e^{i\theta \text{sign}(v)} F(v) \quad (4)$$

where θ is chosen to maximise the peak in $h(t, \theta)$. For $\theta = \phi$ this will restore the phase of the original signal (cf. equation 3), but this is not generally the case. We can re-express the reconstructed signal as

$$H(v, \theta) = F(v) \cos \theta + G(v) \sin \theta \quad (5)$$

where

$$G(v) = i \text{sign}(v) F(v). \quad (6)$$

Equivalently to equation 5, in the time domain,

$$h(t, \theta) = f(t) \cos \theta + g(t) \sin \theta. \quad (7)$$

Since θ is chosen to maximise $h(t, \theta)$, we know that

$$\frac{\partial}{\partial \theta} h(t, \theta) = 0, \quad (8)$$

hence

$$\frac{g(t)}{f(t)} = \tan \theta \quad (9)$$

and $h(t, \theta)$ from equation 7 can be rewritten without θ -dependence as

$$h(t) = \sqrt{f(t)^2 + g(t)^2}, \quad (10)$$

where the choice of the positive square root reflects the choice of a positive peak in $h(t)$. We can express $g(t)$ by using the convolution theorem to find the time-domain equivalent to equation 6,

$$g(t) = \frac{1}{\pi t} * f(t), \quad (11)$$

which is the Hilbert transform of $f(t)$. This makes $h(t)$ the same as the signal envelope defined and extensively discussed by Longuet-Higgins (1957); here we will consider only its excursion rate, in sections 3.3 and 3.4.

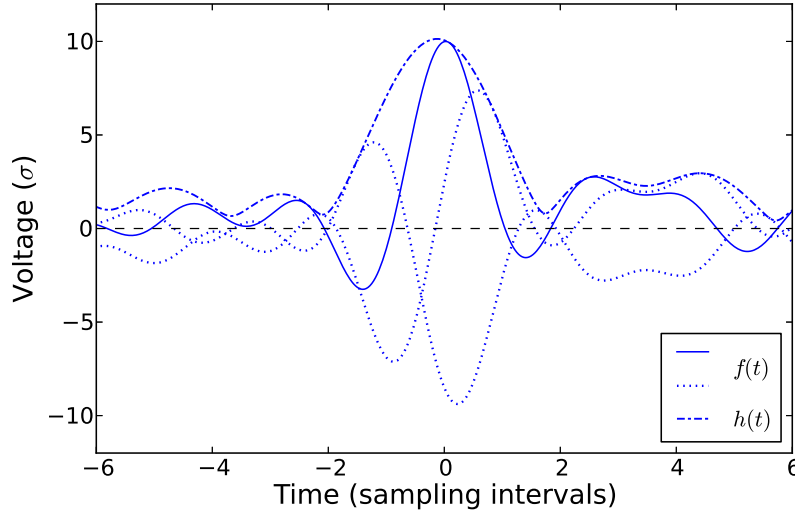


Fig. 1 A coherent, zero-phase, flat-spectrum pulse (i.e. a sinc function) against a background of Gaussian random noise with the same spectrum (solid). Adding a variable phase ϕ to the signal, as per equation 3, results in a range of possible pulse shapes, some of which are shown (dotted), with a decreased peak amplitude. These represent different possibilities for the signal $f(t)$ from section 2.2. The signal envelope $h(t)$ (dash-dotted), from equation 10, recovers the full amplitude of the pulse.

3 Excursion rates

The significance of an event detected in a stochastic signal is dependent upon the expected rate of false detections, given the characteristics of the signal. Here we consider the case of pulse-like events, involving an excursion beyond some high threshold, in a Gaussian random signal with an arbitrary spectrum, and find the expected rate of such excursions.

We give the excursion rate for a raw digital signal (i.e. after sampling, but before further digital processing) in section 3.1, and after perfect interpolation (i.e. for a continuous signal) in section 3.2. These results are known, but the working provides a useful foundation for sections 3.3 and 3.4, in which we derive the equivalent results for the signal envelope (as described in section 2.2).

The excursion rates given here are one-sided, beyond a positive threshold only, following the convention of Rice (1944, 1945). For a Gaussian signal, the two-sided excursion rates will be exactly twice these values. The signal envelope is always positive, so for it the distinction is unimportant.

3.1 Raw signal

Immediately after sampling, a signal such as the measured voltage in a radio receiver consists of a series of Gaussian samples $f(t_i)$ with probability density

$$P(f) = \frac{1}{\sigma\sqrt{2\pi}} e^{-f^2/2\sigma^2} \quad (12)$$

where σ is the standard deviation or root mean square (RMS) signal amplitude, and the mean is assumed to be zero. An excursion in the signal consists of one or more consecutive samples above some threshold f_{thr} . If f_{thr} is sufficiently high that it is rare for two consecutive samples to exceed it, the probability that a particular sample exceeds the threshold gives us the expected rate of such excursions,

$$\begin{aligned} \text{excursion rate} &= \int_{f_{\text{thr}}}^{\infty} df P(f) \\ &= \frac{1}{2} - \frac{1}{2} \operatorname{erf}\left(\frac{f_{\text{thr}}}{\sigma\sqrt{2}}\right) \end{aligned} \quad (13)$$

in units of the sampling rate, with erf being the standard Gauss error function. However, if the signal samples are strongly correlated in time — as for an oversampled signal, or after interpolation — a single excursion will typically involve a number of consecutive samples, so this approach is insufficient.

3.2 Interpolated signal

After interpolation to produce a continuous signal, the excursion rate is found by means of Rice's formula, which gives the rate of upward crossings — and, hence, distinct excursions — over a threshold f_{thr} for a signal $f = f(t)$, as

$$\text{excursion rate} = \int_0^{\infty} df' f' P(f_{\text{thr}}, f') \quad (14)$$

where f' is the time derivative of f , and $P(f, f')$ is their joint probability distribution. An early form of this result appears in Rice (1945, eqn 3.3-5); refer to Rainal (1988) for a detailed discussion. $P(f, f')$ can be rewritten as

$$P(f, f') = P(f) P(f'; f) \quad (15)$$

where $P(f)$ is the distribution of f (from equation 12) and $P(f'; f)$ is the distribution of f' conditional upon the value of f . To find this latter distribution, we consider points at closely separated times t and $t + dt$. If $f(t)$ is a correlated Gaussian signal, then the corresponding values f and $f + df$ follow a joint normal distribution (see e.g. Kenney (1939, ch. 4)): given the value of f , $f + df$ is normally distributed with expected value and variance

$$\begin{aligned} \mathbb{E}[f + df] &= f \rho(dt) \\ \operatorname{Var}[f + df] &= \sigma^2 (1 - \rho(dt)^2) \end{aligned} \quad (16)$$

where $\rho(dt)$ is the value of the autocorrelation function (Fourier transform of the power spectrum, normalised to $\rho(0) = 1$) at separation dt . The slope $f' = df/dt$ is then also normally distributed, with expected value and variance

$$\begin{aligned} E[f'] &= \frac{E[f + df] - f}{dt} \\ \text{Var}[f'] &= \frac{\text{Var}[f + df]}{dt^2}. \end{aligned} \quad (17)$$

Taking the limit $dt \rightarrow 0$ and the expressions from equation 16, and noting that $\rho'(0) = 0$, we obtain

$$\begin{aligned} E[f'] &= 0 \\ \text{Var}[f'] &= -\sigma^2 \rho''(0) \end{aligned} \quad (18)$$

which represents the distribution

$$\begin{aligned} P(f'; f) &= P(f') \quad \text{irrespective of } f \\ &= \frac{1}{\zeta \sqrt{2\pi}} e^{-f'^2/2\zeta^2} \end{aligned} \quad (19)$$

where

$$\begin{aligned} \zeta &= \sqrt{\text{Var}[f']} \\ &= \sigma \sqrt{-\rho''(0)}. \end{aligned} \quad (20)$$

With the expressions for $P(f)$ and $P(f'; f)$ from equations 12 and 19 respectively, equation 14 gives us

$$\text{excursion rate} = \frac{1}{2\pi} \sqrt{-\rho''(0)} e^{-f_{\text{thr}}^2/2\sigma^2} \quad (21)$$

as found by Rice (1945, eqns 3.3-11 & 3.6-11). For discrete data, as in section 5 of this work, $\rho(t)$ can be obtained numerically from the measured power spectrum.

3.2.1 White noise

When the power spectrum is known analytically, equation 21 can be simplified. Consider the special case of white noise, with constant power spectral intensity from frequency ν_a to ν_b , i.e.

$$I(\nu) = \begin{cases} 1 & \nu_a < |\nu| < \nu_b \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

This function has the Fourier transform

$$\mathcal{F}\{I\}(t) = 2\nu_b \text{sinc}(2\nu_b t) - 2\nu_a \text{sinc}(2\nu_a t) \quad (23)$$

where $\text{sinc}(x) = \sin(\pi x)/(\pi x)$. Normalising this, we obtain the autocorrelation function

$$\rho(t) = \frac{\nu_b \text{sinc}(2\nu_b t) - \nu_a \text{sinc}(2\nu_a t)}{\nu_b - \nu_a}. \quad (24)$$

The second derivative of this function at $t = 0$ is

$$\rho''(0) = -\frac{4}{3}\pi^2 \frac{v_a^3 - v_b^3}{v_a - v_b}, \quad (25)$$

which we insert into equation 21 to obtain

$$\text{excursion rate} = \sqrt{\frac{1}{3} \frac{v_a^3 - v_b^3}{v_a - v_b}} e^{-f_{\text{thr}}^2/2\sigma^2}, \quad (26)$$

matching the result found for this case by Rice (1945, eqns 3.3-12 & 3.6-11).

3.3 Signal envelope

The signal envelope $h(t)$, per equation 10, is a function of the signal $f(t)$ and its Hilbert transform $g(t)$. We take both $f = f(t)$ and $g = g(t)$ to be normally distributed with standard deviation σ . The distribution of $h = h(t)$ depends on the degree of correlation between them. Given the value of $f(t)$, the expected value of the signal at other times is

$$\mathbb{E}[f(t - \tau)] = f(t) \rho(\tau) \quad (27)$$

where $\rho(\tau)$ is the autocorrelation function as in section 3.2. The expected value of g , from equation 11, is then

$$\begin{aligned} \mathbb{E}[g] &= \int d\tau \frac{1}{\pi\tau} \mathbb{E}[f(t - \tau)] \\ &= \int d\tau \frac{1}{\pi\tau} f(t) \rho(\tau) \\ &= 0 \end{aligned} \quad \text{as } \frac{1}{\pi\tau} \text{ is odd, and } \rho(\tau) \text{ is even.} \quad (28)$$

Therefore there is no correlation between f and g ; and, from equation 10, h follows a Rayleigh distribution

$$P(h) = \frac{h}{\sigma^2} e^{-h^2/2\sigma^2} \quad (29)$$

as found by Longuet-Higgins (1957). If consecutive values $h(t_i)$ are not significantly correlated, and we choose a threshold h_{thr} sufficiently high that consecutive values are unlikely to exceed it (as in section 3.1), this gives us the excursion rate of the signal envelope:

$$\begin{aligned} \text{excursion rate} &= \int_{h_{\text{thr}}}^{\infty} dh P(h) \\ &= e^{-h_{\text{thr}}^2/2\sigma^2} \end{aligned} \quad (30)$$

in units of the sampling rate.

3.4 Interpolated signal envelope

To find the excursion rate for the interpolated signal envelope h , we apply the approach from section 3.2, which requires us to find the distribution of its time derivative h' . From equation 10,

$$h' = \frac{1}{h} (f f' + g g'). \quad (31)$$

The distribution of this value depends on the correlation between its components. We know from section 3.2 that f' is normally distributed with mean 0 and variance ς^2 , and uncorrelated with f ; the same applies to g' with respect to g . We also know from section 3.3 that f and g are uncorrelated. We now consider the correlation between f and g' , the latter of which is found as

$$\begin{aligned} g'(t) &= \frac{d}{dt} \left(\frac{1}{\pi t} * f(t) \right) && \text{from equation 11} \\ &= -\frac{1}{\pi t^2} * f(t). \end{aligned} \quad (32)$$

Hence the conditional expected value of g' , given f , is

$$\begin{aligned} E[g'] &= - \int d\tau \frac{1}{\pi \tau^2} E[f(t - \tau)] \\ &= - \int d\tau \frac{1}{\pi \tau^2} f(t) \rho(\tau) && \text{from equation 27} \\ &= \alpha \frac{\varsigma}{\sigma} f \end{aligned} \quad (33)$$

where

$$\begin{aligned} \alpha &= -\frac{\sigma}{\varsigma} \int d\tau \frac{1}{\pi \tau^2} \rho(\tau) \\ &= \frac{-1}{\sqrt{-\rho''(0)}} \int d\tau \frac{1}{\pi \tau^2} \rho(\tau) && \text{from equation 20} \end{aligned} \quad (34)$$

is the coefficient of correlation between f and g' . The variance of g' is then

$$\text{Var}[g'] = \varsigma^2 (1 - \alpha^2) \quad (35)$$

conditional upon the value of f .

We consider the correlation between g and f' in a similar manner. First we note, from equation 11, that we can obtain f via the inverse Hilbert transform

$$f(t) = -\frac{1}{\pi t} * g(t). \quad (36)$$

Hence

$$\begin{aligned} f'(t) &= \frac{d}{dt} \left(-\frac{1}{\pi t} * g(t) \right) \\ &= \frac{1}{\pi t^2} * g(t) \end{aligned} \quad (37)$$

and the conditional expected value, given g , is

$$\mathbb{E}[f'] = -\alpha \frac{\zeta}{\sigma} g. \quad (38)$$

The conditional variance, $\text{Var}[f']$, is identical to $\text{Var}[g']$ in equation 35, as it is insensitive to the sign of α .

We know that h' , given f and g , is the sum of two independent normal variables (corresponding to the terms in equation 31), and hence normally distributed itself. We find its expected value and variance:

$$\begin{aligned} \mathbb{E}[h'] &= \frac{1}{h} (f \mathbb{E}[f'] + g \mathbb{E}[g']) \\ &= 0 \end{aligned} \quad \text{from equations 33 and 38} \quad (39)$$

$$\begin{aligned} \text{Var}[h'] &= \frac{1}{h^2} (f^2 \text{Var}[f'] + g^2 \text{Var}[g']) \\ &= \frac{f^2 + g^2}{h^2} \text{Var}[f'] \quad \text{as } \text{Var}[f'] = \text{Var}[g'] \\ &= \zeta^2 (1 - \alpha^2) \end{aligned} \quad \text{from equations 10 and 35.} \quad (40)$$

The distribution of h' is thus

$$\begin{aligned} P(h'; h) &= P(h') \quad \text{irrespective of } h \\ &= \frac{1}{\zeta_h \sqrt{2\pi}} e^{-h'^2 / 2\zeta_h^2} \end{aligned} \quad (41)$$

where

$$\begin{aligned} \zeta_h &= \sqrt{\text{Var}[h']} \\ &= \sigma \sqrt{-\rho''(0)} \sqrt{1 - \alpha^2} \end{aligned} \quad \text{from equations 20 and 40.} \quad (42)$$

With this distribution, and the distribution $P(h)$ from equation 29, we can determine the excursion rate beyond a threshold h_{thr} from equation 14, finding that

$$\text{excursion rate} = \frac{1}{\sqrt{2\pi}} \sqrt{-\rho''(0) - \left(\int d\tau \frac{1}{\pi\tau^2} \rho(\tau) \right)^2} \frac{h_{\text{thr}}}{\sigma} e^{-h_{\text{thr}}^2 / 2\sigma^2}. \quad (43)$$

If the integral in this expression is to be evaluated with numerical data, with values of $\rho(t)$ calculated for discrete time intervals, note that the analytic result

$$\int_{-\varepsilon}^{\varepsilon} d\tau \frac{1}{\pi\tau^2} \rho(\tau) \simeq -\frac{2}{\pi\varepsilon} \rho(0) \quad (44)$$

should be used for the interval $-\varepsilon < \tau < \varepsilon$ to avoid the singular behaviour at $\tau = 0$.

3.4.1 White noise

Here we consider white noise from frequency ν_a to ν_b as in section 3.2.1, and derive the excursion rate for the signal envelope in this case. The autocorrelation function for white noise is given by equation 24, which allows us to evaluate the integral in equation 43 as

$$\int d\tau \frac{1}{\pi\tau^2} \rho(\tau) = -\pi(\nu_a + \nu_b). \quad (45)$$

Also incorporating $\rho''(0)$ from equation 25, we can evaluate the excursion rate from equation 43 to be

$$\text{excursion rate} = \sqrt{\frac{\pi}{6}} |\nu_a - \nu_b| \frac{h_{\text{thr}}}{\sigma} e^{-h_{\text{thr}}^2/2\sigma^2}. \quad (46)$$

4 Monte Carlo simulation

The results from section 3 can be tested through Monte Carlo simulation: generating random noise data and comparing the distribution of peak heights (for the signal and its envelope, with and without interpolation) with the analytic predictions. This also provides a helpful illustration of the effects of interpolation and the phase search represented by the signal envelope on the excursion rate.

4.1 Uncorrelated noise

The simplest case to test through simulation is that of uncorrelated noise. We generated 100,000 arrays, each containing 4,096 independent normally-distributed random values. For each array, we performed the following steps:

1. Find the maximum value in the array.
2. Perform interpolation on the array, and find the maximum interpolated value.
3. Take the norm of the array and its Hilbert transform to find the signal envelope, as in equation 10. Find the maximum value in the resulting array.
4. Repeat step 3 with an interpolated version of the array, and find the maximum value.

The interpolation in steps 2 and 4 was 32-fold, with 31 interpolated values between each pair of raw values. This level of precision was chosen as it resulted in maximum reconstructed values within 0.1% of those resulting from finer searches, while being relatively inexpensive to compute.

The number of excursions beyond some threshold, under each of the conditions above (raw values; interpolation; signal envelope; interpolated envelope), can be considered to be the number of arrays containing a peak height above that threshold. This leads to an underestimate of the excursion rate for low thresholds, for which multiple excursions within a single 4,096-value array are common; but it is accurate for high thresholds, for which this is rare.

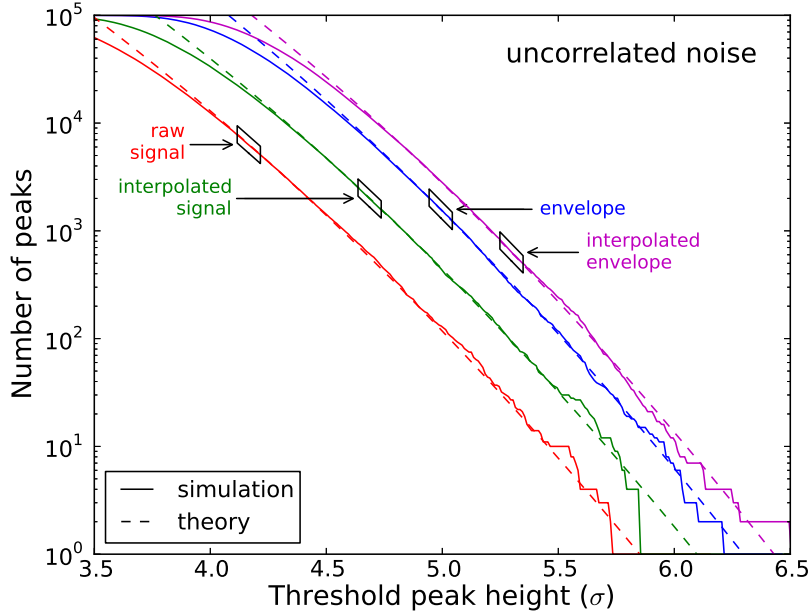


Fig. 2 Cumulative histograms (solid) of the peak height in an array for white-spectrum Monte Carlo data, for the signal and its envelope, with and without interpolation, compared with analytic predictions (dashed). The divergence at low thresholds is due to arrays containing multiple peaks, which are counted only once. The divergence at high thresholds is due to small-number statistics.

As the values are uncorrelated, the expected excursion rates at high thresholds for the uninterpolated data are given by equation 13 for the raw signal and equation 30 for the signal envelope. Uncorrelated values also result in a white spectrum, with a constant spectral intensity up to an implied maximum frequency of 0.5 cycles per sampling interval, allowing the use of equations 26 and 46 to predict the excursion rates for interpolated data, for the signal and its envelope respectively. In figure 2 we compare these excursion rates, in terms of the total expected number of excursions, with cumulative histograms of the peak height in an array, finding them in good agreement.

4.2 Correlated noise

To test the dependence of the excursion rate on the autocorrelation of the noise, we performed another Monte Carlo simulation with a different spectrum. Arrays of random values were generated with a Gaussian power spectrum

$$I(\nu) \propto e^{-\nu^2/2\sigma_\nu^2} \quad |\nu| < \nu_s/2 \quad (47)$$

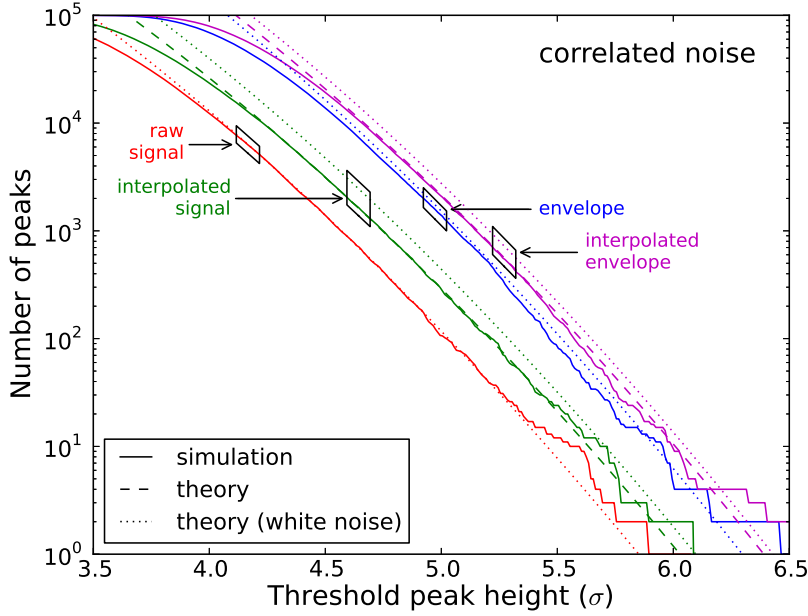


Fig. 3 Cumulative histograms (solid) of the peak height in an array for Gaussian-spectrum Monte Carlo data, for the signal and its envelope, with and without interpolation, compared with analytic predictions (dashed). Divergence at low and high thresholds, as in figure 2, is due to under-counting of peaks and small-number statistics respectively. Analytic predictions for the case of uncorrelated noise (dotted) are shown for comparison.

where $\sigma_v = v_s/5$ is defined in terms of the sampling rate v_s . Peak heights for the signal and its envelope, with and without interpolation, were then found through the same procedure as in section 4.1.

As the values in this case are correlated, the excursion rates for interpolated data must be found with equation 21 for the signal and equation 43 for its envelope, in terms of the autocorrelation function $\rho(t)$. This function can be derived analytically from the power spectrum in equation 47, but was instead obtained numerically in order to validate the procedure to be used in section 5. Unlike the uncorrelated data in section 4.1, excursion rates for correlated uninterpolated data are not expected to be given exactly by equations 13 and 30.

Cumulative histograms of peak heights are compared with these predictions in figure 3. The histograms for interpolated data agree closely with the predictions, significantly lower than the excursion rates expected for uncorrelated noise.

5 Experimental test with the Parkes Bedlam backend

Bedlam is a digital backend for the Parkes radio telescope, designed and built to perform transient pulse detection for the LUNASKA project (Bray et al., 2012). We

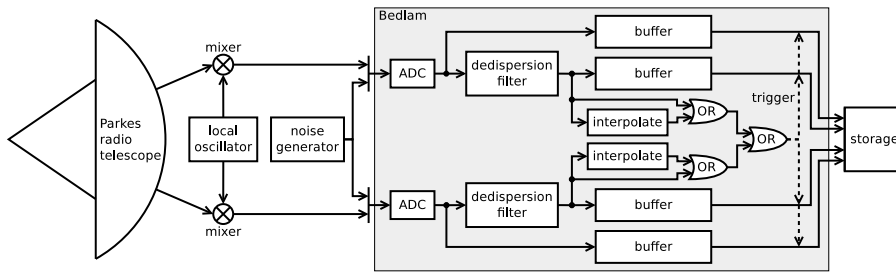


Fig. 4 Signal path for two active channels in the Parkes Bedlam backend (see section 5.1 for details). In normal operation (outer signal paths) it receives input from the Parkes radio telescope; in this experiment, it was instead connected to a noise generator (centre). A trigger condition (dashed) causes the contents of the buffers to be stored.

performed an experiment in which Bedlam processed data from a noise source, the results of which allow us to test the performance of the hardware and the results from section 3.

5.1 The Parkes Bedlam backend

In normal operation, Bedlam is connected to the Parkes 21 cm multibeam receiver (Staveley-Smith et al., 1996). It has eight inputs, typically receiving the voltage signal from both polarisations for four of this receiver’s thirteen beams. The receiver amplifies the signal and downconverts it to a lower frequency (see section 2.1); it then enters Bedlam where it is digitised by analog-to-digital converters (ADCs) at 1,024 Msamples/s (corresponding to 512 MHz of bandwidth) and 8 bits of precision, and subjected to an adjustable dedispersion filter to compensate for ionospheric dispersion. Buffers of the raw and dedispersed signals are kept, with an adjustable size up to 8 μ s.

The dedispersed data are then interpolated. Full reconstruction of a Nyquist-sampled signal is achieved by convolving with a sinc function; interpolation here is performed as a convolution with a six-point approximation to a sinc function (at ± 0.5 , ± 1.5 and ± 2.5 sampling intervals), producing a single interpolated point between every pair of original samples. If a value is found over an adjustable threshold, in either the original or interpolated values, it triggers the permanent storage of both raw and dedispersed buffers for all input channels. Anticoincidence logic can be applied between channels to prevent triggering from simultaneous above-threshold values in multiple channels, but was inactive for this experiment.

In this experiment, all eight channels were connected to the same noise source. However, thresholds were set such that only two of these were able to trigger buffer storage. The signal path for these channels is shown in figure 4.

5.2 Experiment

The Bedlam backend was operated with pure noise input for 478 hours, starting on 23 September 2010. The noise source was a reverse-biased diode in breakdown, with

a passive eight-way power-splitter to connect it to all eight input channels. In this time, there were 902,115 peaks in the noise signal sufficient to trigger storage of the buffered data, with the buffer size set to $4\ \mu\text{s}$ (4,096 samples). During the 51 ms period taken to store a single set of buffers of this length, the system is unable to respond to any further triggers; this dead time amounts to 2.7% of the experimental period. The dedispersion filter was set to its minimal setting, equivalent to $\sim 2\ \text{ns}$ differential delay across the band of the noise.

The RMS noise amplitude was typically ~ 12 analog-to-digital units (ADU), compared to a maximum of +127 ADU for an 8-bit signed integer. The temperature of the noise source was not controlled, so the gain varied over the course of the experiment, resulting in $\sim 5\%$ slow variation in the RMS. More dramatic fluctuations occurred during a 8-hour period partway through the experiment, possibly due to a problem with the power supply, so this interval was excluded. After also excluding dead time, the remaining effective duration of the experiment is 458 hours.

Typical stored buffers consisted mostly of pure noise, with a single extreme value that exceeded the trigger threshold in the centre of the buffer, as expected from Gaussian noise with a relatively wide band. However, there were two atypical events which exhibited increased signal power for an extended period, over 100 ns (see figure 5). These events are not consistent with pure Gaussian noise; their origin is unknown, but more such events were found during the period excluded due to RMS fluctuations, which suggests that they may be related phenomena. These two events were excluded from further analysis.

A minor bug resulted in spurious recorded values at the start and end of each buffer, so these were excluded from the analysis; the triggering value, at the centre of the buffer, is unaffected by this. Peaks in the spectrum at 0 and 512 MHz — the bottom and top of the band — were removed, the former being a direct-current offset and the latter probably resulting from internal interference from the digital hardware, at a harmonic of its clock speed of 256 MHz. Lesser interference at harmonic and subharmonic frequencies was considered insignificant and not removed, but is just visible in the noise spectrum shown in figure 6.

5.3 Excursion rates

Excursion rates for the signal and its envelope, with and without interpolation, were found as in section 4 for post-dedispersion buffers for a single channel of Bedlam. Although the signal in this case does not undergo frequency mixing, unlike normal operation of the Parkes radio telescope (see figure 4), random noise has no preferred phase and hence displays the same characteristics with or without this step. For the same reason, the dedispersion filter has no effect on the characteristics of the signal other than to change its spectrum. The spectrum after filtering is shown in figure 6 and the derived autocorrelation, for determining the expected excursion rates, is shown in figure 7.

The significance of the peak in each buffer was calculated against a value of σ equal to the combined RMS of all buffers within a moving 10-minute window, to reduce the uncertainty associated with the limited number of samples in a single buffer.

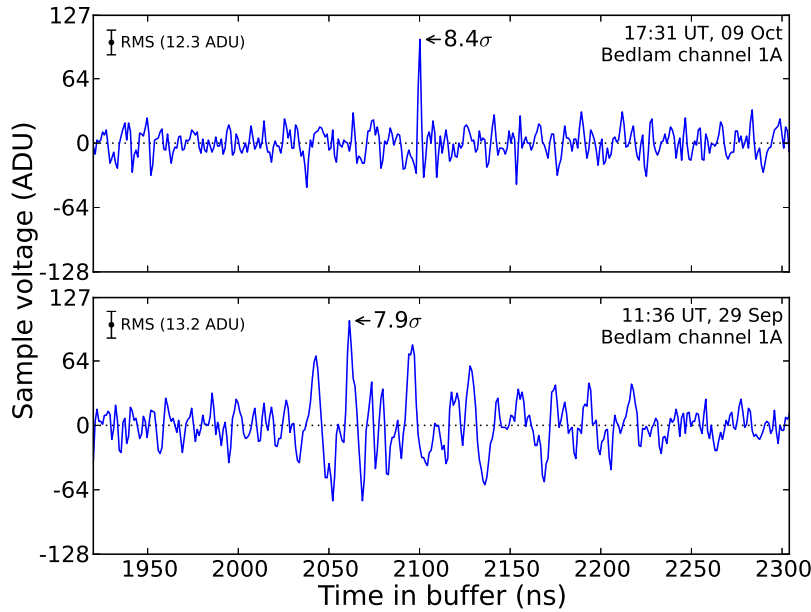


Fig. 5 The two most significant peaks, after dedispersion, recorded over the course of the experiment. The upper peak displays the narrow profile typical of rare fluctuations in Gaussian noise. The lower peak includes an extended elevation in signal power, possibly indicative of an instrumental problem, and is one of two such peaks excluded from the analysis (see text). In both cases, similar peaks are seen in the other channels.

The central portion of each buffer, containing the triggering value, was excluded from the RMS calculation to avoid bias. The RMS variation over a 10-minute period was $\sim 0.5\%$, resulting in a corresponding uncertainty in the peak height.

Quantisation error in the ADCs of ± 0.5 ADU is equivalent to 0.5% at 8σ . Summed in quadrature with the uncertainty from RMS variation, above, this gives a 0.7% uncertainty in the height of a single peak. This has a negligible direct effect on the measured excursion rates for low threshold values, as it is averaged over a large number of excursions; it applies in full only to the single highest peak. However, it indirectly causes a systematic effect: due to random increases or decreases in the peak height, the rate of excursions beyond a particular threshold will be a weighted average of the excursion rates that would occur at lower or higher thresholds in the absence of this variation. This effect is equivalent to convolving the excursion rate as a function of threshold with the random error distribution. Making a Gaussian approximation to the latter, we find that 0.7% random error leads to a $\sim 0.2\%$ increase in the threshold at which a given excursion rate occurs.

Figure 8 shows the measured excursion rates compared to the predictions from equations 21 and 43, modified by the above effect. As in section 4.2 (figure 3), predictions for the case of uncorrelated white noise are also shown. The results are in reasonable agreement: the difference between the excursion rates for the signal and

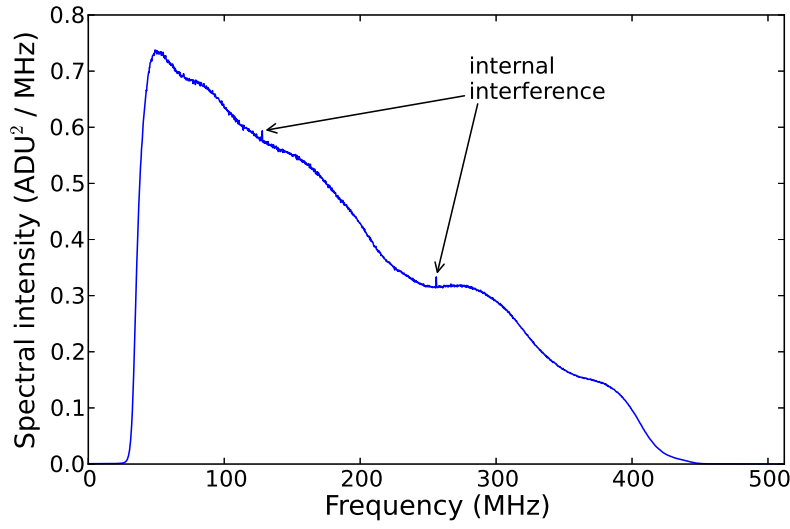


Fig. 6 Power spectrum of the recorded dedispersion-filtered noise data. The peaks at 128 and 256 MHz (marked) are internal interference from the digital signal-processing hardware.

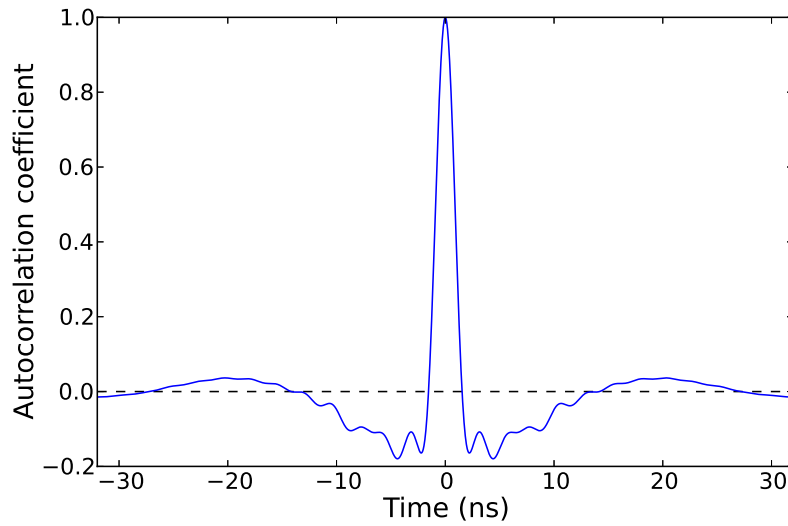


Fig. 7 Autocorrelation function $\rho(t)$ for the recorded noise data around $t = 0$, derived from the power spectrum in figure 6.

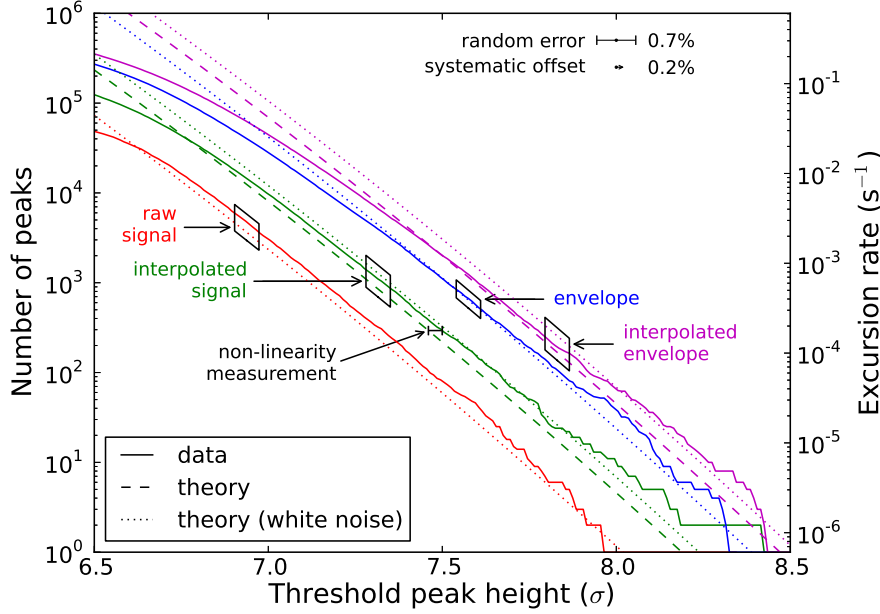


Fig. 8 Measured excursion rates (solid) of the peak height in each buffer for experimental data with the Parkes Bedlam backend, for the signal and its envelope, with and without interpolation, compared with analytic predictions (dashed). At low thresholds, there is a deficit in the excursion rates due to incomplete triggering. At high thresholds, there is divergence due to random error and small-number statistics. The systematic offset associated with the random error (see text) has already been applied to the predicted curves; we ascribe the remaining consistent excess at intermediate thresholds to non-linearity, which is measured as shown and discussed in section 5.4. Analytic predictions for the case of uncorrelated noise (dotted) are shown for comparison.

for its envelope is clearly visible, although the increased error in the experimental data relative to the Monte Carlo simulation makes it impossible to clearly discriminate between the predictions for uncorrelated noise and for the actual spectrum.

The deficit in the excursion rate at low thresholds is not caused by multiple excursions in a single array, as in the Monte Carlo data, but by the Bedlam backend failing to trigger on some lower-amplitude peaks. As the triggering algorithm acts only on the signal (rather than its envelope), and performs only partial interpolation, a peak which is of low amplitude in the raw data but high amplitude after interpolation or in the envelope will be missed; consequently, the deficit extends to higher thresholds for interpolated and envelope data. There is divergence at high thresholds from small-number statistics, as for the Monte Carlo data, but with an additional contribution from the random error described above.

The remaining discrepancy in figure 8 is the small systematic excess at intermediate thresholds, most clearly visible for the raw and interpolated data. Measuring this excess relative to the theoretical excursion rate for interpolated data at 7.5σ , as a compromise between uncertainty from incomplete triggering (at lower thresholds)

and from small-number statistics (at higher thresholds), this is equivalent to a 0.4% offset in the threshold. The number of excursions beyond this threshold is 278, which is sufficient to make the averaged random error in the peak height negligible, but there is a remaining effect from small-number statistics: the Poisson error of $\pm 1/\sqrt{278}$ in the number of excursions results in a $\pm 0.1\%$ uncertainty in the measurement of this offset. We discuss the cause of this offset in the next section.

5.4 Non-linearity

In an ideal ADC, there is a direct linear relationship between the thresholds in the analog input voltage required to produce particular digital output codes, and the values of those output codes (in ADU). In practice, there are minor offsets between these thresholds and their ideal values. The maximum such offset for a given ADC is its integral non-linearity (INL); for the hardware used in this experiment, this is advertised as 0.4 ADU (e2v, 2010). This non-linearity affects measured peak heights, and hence excursion rates, through two different mechanisms.

The first effect is a direct error in the measurement of the peak height, with a magnitude less than or equal to the INL. For this experiment, an INL of 0.4 ADU leads to a possible 0.4% error for peaks at 7.5σ . The direction and magnitude of the error may vary across the voltage range of the ADC, controlled by the difference in the non-linearity between adjacent voltage thresholds (the differential non-linearity, or DNL). In general, we expect the error for peaks in the positive direction (around +100 ADU) to be uncorrelated with the error for peaks in the negative direction (around -100 ADU).

The second effect arises from mismeasurement of the RMS, which leads to an error in the normalised peak height (measured in σ). This effect is dominated by the non-linearity in voltage thresholds within $1-2\sigma$ (12–24 ADU) of zero voltage, within which most sample values lie. The magnitude of this effect is difficult to predict, as it depends strongly on the variation of the non-linearity across this range, but we can place an approximate upper bound of the INL (0.4 ADU) divided by the RMS, or $\sim 3\%$, corresponding to minimal variation. Because the same RMS value is used to normalise both positive and negative peak heights, we expect this error to apply to measured excursion rates in both directions.

The offset of $0.4 \pm 0.1\%$ in the threshold peak height for positive excursions in the first input channel of Bedlam, found in section 5.3, is within the possible range of the above effects. Repeating the analysis for negative excursions, we find an offset of $0.9 \pm 0.1\%$; the difference between these is explained entirely by the $\pm 0.4\%$ error arising directly from the peak height measurement. Note that these offsets are both for the interpolated signal, rather than the envelope: as the envelope is defined to be positive, it neglects the distinction between positive and negative excursions.

Similar results were obtained for all input channels of Bedlam, each of which passes through a separate ADC. The results are shown in table 1, and are all within the $\pm 3\%$ range expected from RMS measurement error. There is a clear correlation between the positive and negative excursion rates for each channel; the two highest and two lowest excesses are associated with channels 2A and 4B respectively. This

channel	non-linearity (%)	
	positive	negative
1A	+0.4	+0.9
1B	+0.3	+1.0
2A	+1.7	+2.5
2B	+1.1	+1.2
3A	+0.5	+0.4
3B	+0.8	+1.0
4A	+0.8	+1.6
4B	-0.4	-0.3

Table 1 Offsets in threshold peak height, ascribed to non-linearity, required to produce the observed excess for rates of positive and negative excursions for interpolated data in all channels; figure 8 shows the measurement for positive excursions in channel 1A. The uncertainty in all measurements is $\pm 0.1\%$. The channel labelling scheme reflects the normal usage of the Bedlam backend, with each channel connected to either the A or B polarisation of one of four beams of the Parkes 21 cm multibeam receiver.

is consistent with the greater part of the common excess originating from error in the RMS measurement, while the difference between the excesses for positive and negative excursions is caused directly by errors in measurement of the peak height. The greatest such difference is 0.8% for channel 2A, which is just compatible with independent errors of $\pm 0.4\%$ in measurement of the peak height, combined with $\pm 0.1\%$ from measurement of the excess in the excursion rate.

6 Conclusion

Experiments which search for rare, transient events against a background of Gaussian noise, with a time-scale comparable to the inverse bandwidth, require a thorough understanding of the noise statistics to evaluate the significance of a possible detection. If the signal undergoes a frequency mixing step, a phase ambiguity is introduced, and a full search for a possible event requires the computation of the signal envelope. We have derived the expected excursion rate from Gaussian noise under these conditions, and confirmed it through simulation and experiment. We have also described and tested Bedlam, a digital backend for the Parkes radio telescope which is designed for such an experiment.

Acknowledgements The Parkes radio telescope is part of the Australia Telescope which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. This research was supported by the Australian Research Council’s Discovery Project funding scheme (project number DP0881006).

References

- EV8AQ160 ADC Datasheet. e2v Semiconductors (Jun 2010)
 Beresinsky, V.S., Zatsepin, G.T.: Cosmic rays at ultra high energies (neutrino?).
 Phys. Lett. B28, 423–424 (Jan 1969)

- Bray, J.D., Ekers, R.D., Roberts, P., Reynolds, J.E., James, C.W., Phillips, C.J., McFadden, R.A., Protheroe, R.J., Aartsen, M., Álvarez-Muñiz, J.: Status and strategies of current LUNASKA lunar Cherenkov observations with the Parkes radio telescope. *Nucl. Instrum. Meth. A*662, Supplement 1(0), S95–S98 (2012), 4th International workshop on Acoustic and Radio EeV Neutrino detection Activities
- Buitink, S., Scholten, O., Bacelar, J., Braun, R., de Bruyn, A.G., Falcke, H., Singh, K., Stappers, B., Strom, R.G., Yahyaoui, R.A.: Constraints on the flux of ultrahigh energy neutrinos from Westerbork Synthesis Radio Telescope observations. *A&A*521, A47+ (Oct 2010)
- Jaeger, T.R., Mutel, R.L., Gayley, K.G.: Project RESUN, a Radio EVLA Search for UHE Neutrinos. *Astropart. Phys.*34, 293–303 (Dec 2010)
- James, C.W., Ekers, R.D., Álvarez-Muñiz, J., Bray, J.D., McFadden, R.A., Phillips, C.J., Protheroe, R.J., Roberts, P.: LUNASKA experiments using the Australia Telescope Compact Array to search for ultrahigh energy neutrinos and develop technology for the lunar Cherenkov technique. *Phys. Rev. D*81(4), 042003+ (Feb 2010)
- Kenney, J.F.: *Mathematics of Statistics*, vol. 2. Chapman & Hall, London, 4 edn. (1939)
- Longuet-Higgins, M.S.: The Statistical Analysis of a Random, Moving Surface. *Royal Society of London Philosophical Transactions Series A* 249, 321–387 (Feb 1957)
- Rainal, A.J.: Origin of Rice’s formula. *IEEE Trans. on Inf. Theor.*34(6), 1383–1387 (Nov 1988)
- Rice, S.O.: Mathematical Analysis of Random Noise. *Bell Systems Tech. J.*23(3), 282–332 (Jul 1944)
- Rice, S.O.: Mathematical Analysis of Random Noise. *Bell Systems Tech. J.*24(1), 46–156 (Jan 1945)
- Staveley-Smith, L., Wilson, W.E., Bird, T.S., Disney, M.J., Ekers, R.D., Freeman, K.C., Haynes, R.F., Sinclair, M.W., Vaile, R.A., Webster, R.L., Wright, A.E.: The Parkes 21 cm multibeam receiver. *PASA*13, 243–248 (Nov 1996)
- Weinreb, S.: A digital spectral analysis technique and its application to radio astronomy. *Tech. Rep.* 412, M.I.T. Research Lab. of Electronics (Aug 1963)